Perbandingan Klasifikasi Tingkat Keganasan *Breast Cancer* Dengan Menggunakan Regresi Logistik Ordinal Dan *Support Vector Machine* (SVM)

Farizi rachman dan Santi Wulan Purnami
Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember
Jl. Arief Rahman Hakim, Surabaya 60111
E-mail: Santi_wp@statistika.its.ac.id

Abstrak— ***Breast Cancer* merupakan jenis kanker yang sangat berbahaya di dunia. Setiap tahun 506.000 penduduk di dunia meninggal akibat *breast cancer*. Indonesia merupakan salah satu negara berkembang dengan penderita *breast cancer* yang banyak. Hal ini dibuktikan dengan data Sistem Informasi Rumah Sakit (SIRS) 2007 yang menunjukkan *breast cancer* menempati urutan pertama pada pasien rawat inap di rumah sakit seluruh Indonesia (16.85%). Tingginya kasus penyakit ini di Indonesia mengharuskan penderita *breast cancer* untuk melakukan pemeriksaan tingkat keganasan *breast cancer* dengan memperhatikan jenis stadium *breast cancer*. Pada penelitian ini dilakukan analisis tingkat keganasan *breast cancer* dengan menggunakan regresi logistik ordinal dan *support vector machine* (SVM). Berdasarkan hasil penelitian dengan metode regresi logistik ordinal, menunjukkan variabel yang berpengaruh terhadap tingkat keganasan breast cancer adalah ukuran tumor dan nodus dengan ketepatan klasifikasi tertinggi 56.60%. Sedangkan ketepatan klasifikasi dengan menggunakan SVM ketepatan klasifikasi tertinggi dengan menggunakan kernel RBF dan polynomial mencapai 98.11%.**

***Kata Kunci*—*Breast Cancer*, Regresi Logistik Ordinal, *Support Vector Machine*, Klasifikasi.**

I. PENDAHULUAN

Breast cancer merupakan jenis kanker yang sangat berbahaya didunia, baik di negara maju atau negara berkembang. Setiap tahun 12 juta orang di seluruh dunia menderita kanker dan 7,6 juta di antaranya meninggal dunia, dari jumlah tersebut 506.000 disebabkan oleh *breast cancer* [1]. Hal ini menunjukkan bahwa *breast cancer* adalah salah satu kanker ganas di dunia. Saat ini 16 % dari semua jenis kanker pada wanita didunia adalah *breast cancer*. Berdasarkan data WHO, 69 % dari kematian *breast cancer* didunia terjadi di negara berkembang [1].

Indonesia merupakan salah satu negara berkembang dengan penderita *breast cancer* yang banyak. Berdasarkan data Sistem Informasi Rumah Sakit (SIRS) tahun 2007, breast cancer menempati urutan pertama pada pasien rawat inap di seluruh RS di Indonesia (16,85%). *Breast cancer* merupakan kanker tertinggi yang diderita perempuan Indonesia yaitu 26 per 100.000 perempuan [2]. Hal ini menunjukkan bahwa *breast cancer* adalah penyakit yang sangat berbahaya di Indonesia. Berdasarkan fakta tersebut dibutuhkan suatu langkah strategis untuk deteksi dini *breast cancer* di Indonesia.

Breast cancer merupakan penyakit yang sangat ganas dan diketahui secara pasti penyebabnya. Tingginya kasus *breast cancer* Indonesia mengharuskan penderita *breast cancer* untuk melakukan pemeriksaan intensif terkait identifikasi penyakitnya. Untuk wanita yang positif terjangkit *breast cancer*, dan sudah melakukan tahap pengobatan, maka deteksi keganasan *breast cancer* secara berkala sangat penting. Salah satu cara untuk mendeteksi tingkat keganasan *breast cancer* adalah dengan prognosis. Prognosis adalah "tebakan terbaik" tim medis dalam menentukan sembuh atau tidaknya pasien dari penyakit *breast cancer*. Prognosis memiliki manfaat membantu memilih terapi yang tepat, memungkinkan komparasi berbagai terapi di antara sejumlah pasien dengan resiko kekambuhan atau morbiditas yang serupa dan meningkatkan pengetahuan tentang *breast cancer* guna mengembangkan strategi-strategi baru untuk penanganannya [3]. Secara umum tingkat keganasan breast cancer diukur dengan memperhatikan stadium penderita *breast cancer* yaitu stadium I, II, III, dan IV.

Berdasarkan uraian tersebut salah satu metode yang bisa digunakan untuk deteksi tingkat keganasan *breast cancer* adalah SVM (*Support Vektor Machine),* metode ini merupakan metode statistik yang bisa digunakan untuk menetukan identifikasi *breast cancer* setelah dilakukan *treatment*, dengan memperhatikan faktor-faktor identifikasi. Penelitian sebelumnnya oleh Maglogiannis dan Zafiropolous (2007) melakukan diagnosis dan prognosis *breast cancer* dengan menggunakan SVM [4]. Hasil penelitian tersebut menunjukkan bahwa ketepatan klasifikasi menggunakan SVM mencapai 97%. Sedangkan pada penelitian lain oleh Chen, Yang, dan Lie (2011) yang melakukan diagnosis *breast cancer* menunjukkan bahwa ketepatan klasifikasi mencapai 99.1% [5]. Sedangkan pada penelitian lain tentang diagnosis *breast cancer* dengan menggunakan SSVM memiliki tingkat akurasi 97.22% [6]. Berdasarkan latar belakang di atas maka penelitian ini akan menganalisis tingkat keganasan *breast cancer* dengan menggunakan metode regresi logistik ordinal dan *Support Vektor Machine* (SVM).

II. LANDASAN TEORI

A. *Model Regresi Logistik Ordinal*

Secara umum regresi logistik ordinal merupakan salah satu metode statistika untuk menganalisis variabel respon yang mempunyai skala data ordinal yang memiliki 3 kategorik atau lebih. Pada regresi logistik ordinal model berupa kumulatif logit model. Sedangkan untuk variabel prediktor yang digunakan berupa data kategorik dan atau kuantitatif. Sifat

ordinal dari respon Y pada model logit ini dituangkan dalam peluang kumulatif sehingga kumulatif logit model merupakan model yang didapat dengan membandingkan peluang kumulatif yaitu peluang kurang dari atau sama dengan kategori respon ke-j pada p variabel prediktor yang dinyatakan dalam vektor $\mathbf{x_i}$ adalah $P(Y \le j \mid \mathbf{xi})$, dengan peluang lebih besar kategori respon ke-j yaitu $P(Y > j \mid \mathbf{xi})$ [3]. Sedangkan untuk nilai variabel $\mathbf{X}=[x_1, x_2, ..., x_p]^T$ dan parameter $\boldsymbol{\beta} = [\beta_1, \beta_2, ..., \beta_p]^T$ dan kumulatif ke-j sebagai berikut:

$$\pi_k(x_c) = P(Y \le j) \quad = \pi_1 + \pi_2 + \ldots + \pi_r \tag{1}$$

Setelah dilakukan transformasi logistik menjadi model regresi logistik (logit) ordinal atau logit kumulatif:

$$Logit[P(Y \le j)] = \left(\beta_{oj} + \sum_{k=1}^{p} \beta_k x_k \right) \tag{2}$$

Dengan nilai β_k untuk k=1,2,…*p* pada setiap model regresi logistik ordinal adalah sama.

Selanjutnya dilakukan penaksiran parameter. Bentuk umum dari fungsi likelihood untuk sampel dengan n independen observasi (y_i , x_i) adalah

$$l(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left[\pi_0(x_i)^{y0i} \pi_1(x_i)^{y1i} \pi_2(x_i)^{y2i} \right] \tag{3}$$

Dengan nilai i=1,2,…n

Persamaan 3 ini dengan mengggunakan iterasi *Newton Rhapshon,* akan didapatkan taksiran parameter.

Untuk pengujian parameter dilakukan dua kali yaitu uji serentak dan uji parsial. Uji serentak digunakan untuk memeriksa kemaknaan koefisien β secara keseluruhan sebagai berikut:

H_0: $\beta_1 = \beta_2 = ... = \beta_p = 0$

H_1: paling sedikit ada satu $\beta_k \neq 0$ dengan k=1,2,..*p*

Statistik ujiyang digunakan adalah statistik uji G atau likelihood ratio

$$G^2 = \frac{-2 \ln\left(\frac{n_2}{n}\right)^{n_2} \left(\frac{n_1}{n}\right)^{n_1} \left(\frac{n_0}{n}\right)^{n_0}}{\prod_{i=1}^{n} (\pi_i)^{y_i} (1-\hat{\pi}_i)^{(1-y_i)}} \tag{4}$$

Daerah penolakan H_0 adalah jika $\boldsymbol{G^2} > \chi^2_{(\alpha,v)}$ dengan db=v. Pada pengujian ini $\boldsymbol{G^2}$ menyebar mengikuti distribusi Chi square dengan derajat bebas p [3].

Uji parsial dilakukan untuk menguji signifikansi parameter terhadap variabel respon. Pengujian signifikansi parameter menggunakan uji Wald [6] dengan menggunakan hipotesis sebagai berikut:

H_0: $\beta_k = 0$

H_1: $\beta_k \neq 0$. Dengan k=1,2,..p

Perhitungan uji Wald adalah sebagai berikut

$$W = \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)} \tag{5}$$

Untuk daerah penolakan H_0 adalah jika $|W| > Z_{\alpha/2}$

Setelah dilakukan uji serentak dan parsial, maka dilakukan uji kesesuaian model denan menggunakan uji devians. Uji hipotesis yang digunakan sebagai berikut:

H_0: $\hat{\pi}_i = y_i$ atau Model sesuai (tidak ada perbedaan yang nyata antara hasil observasi dengan kemungkinan hasil prediksi model)

H_1: $\hat{\pi}_i \neq y_i$ atau Model tidak sesuai (ada perbedaan yang nyata antara hasil observasi dengan kemungkinan hasil prediksi model).

Statistik uji diatas adalah:

$$D = -2\sum_{i=1}^{n} y_i \ln\left(\frac{\hat{\pi}_i}{y_i}\right) + (1-y_{ij}) \ln\left(\frac{1-\pi_{\hat{i}}}{1-y_i}\right) \tag{6}$$

Dengan $\hat{\pi}_i = \hat{\pi}(x_i)$ merupakan peluang observasi ke-*i* pada kategori ke-*j*. Derajat bebas uji ini adalah (J-(p+1)) dimana J merupakan jumlah kovariat dan *p* merupakan jumlah variabel predictor. Interpretasi regresi logistik ordinal dapat dijelaskan dengan odd ratio. Nilai odd ratio yaitu nilai yang menunjukan perbandingan tingkat kecenderungan dari dua kategori dalam satu variabel prediktor dengan salah satu kategorinya dijadikan pembanding atau kategori dasar.

B. Support Vector Machine

Support Vektor Machine (SVM) melakukan suatu teknik untuk menemukan fungsi pemisah yang bisa memisahkan dua set data dari dua kelas yang berbeda [7]. Metode ini merupakan metode *learning machine* yang bekerja atas prinsip *Sructural Risk Minimazation (SRM)* dengan tujuan menemukan *hyperplane* terbaik yang memisahkan dua buah kelas pada *input space* [8].

Pada dasarnya SVM bekerja dengan prinsip *linier clasifier*, kemudian dikembangkan untuk dapat bekerja pada kasus non linear dengan menggunakan konsep kernel pada ruang kerja berdimesi tinggi [8]. Pada klasifikasi linear SVM dibagi menjadi 2 jenis yaitu *separable* dan *nonseparable.* Misalkan X memiliki pola tertentu, yaitu apabila x_i termasuk kedalam kelas maka x_i diberi label (target) y_i =+1 dan y_i = -1. Untuk itu, label masing-masing dinotasikan $y_i \in \{-1,+1\}$, i=1,2,..l. Sehingga data berupa pasangan (x_1,y_1), (x_2,y_2),..,(x_l,y_l). Kumpulan data pasangan ini merupakan data bagi SVM. *Support Vector Machine* (SVM) bisa menentukan pola generalisasi dari $x \in X$.

Pada dasarnya SVM merupakan metode untuk melakukan klasifikasi himpunan *vektor training* dari dua kelas (x_1,y_1), (x_2,y_2),..,(x_l,y_l), dengan $\mathbf{x} \in R^n$, $y \in \{-1,1\}$.

Pada pemisahan *hyperplane* dengan bentuk *canonical* harus memenuhi constraint atau bisa disebut fungsi kendala

$$y_i[(\mathbf{w} \cdot \mathbf{x}_i) + b] \ge 1, i = 1, ...l \tag{7}$$

Hyperplane yang memisahkan data harus meminimalkan

$$\boldsymbol{\varphi}(w) = \frac{1}{2}\|\mathbf{w}\|^2 \tag{8}$$

Untuk optimasi pada persamaan 8, fungsi *lagrange* yang digunakan adalah :

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{l} \alpha_i \{[(\mathbf{x}_i \cdot \mathbf{w}) + b] y_i - 1\} \tag{9}$$

Pada persamaan tersebut, nilai α_i adalah pengganda fungsi *lagrange*. Solusi dari fungsi Lagrange ini dapat diperoleh dengan meminimalkan L terhadap *primal variables* dan memaksimalkan L terhadap (*dual variables*) dan penyelesaian dari persamaan sebagai berikut:

$$\hat{\alpha} = \arg\min_{\alpha} \frac{1}{2}\sum_{i,j=1}^{l}\alpha_i\alpha_j y_i y_j \left(\mathbf{x}_i.\mathbf{x}_j\right) - \sum_{i=1}^{l}\alpha_i \quad (10)$$

Sehingga persamaan klasifikasi menggunakan persamaan :

$$f(x) = sign\left(\mathbf{w}^T \cdot \mathbf{x} + \hat{b}\right) \quad (11)$$

Untuk mengatasi *misclassification*, formulasi yang telah dilakukan sebelumnnya, akan diperluas sehingga dapat digunakan data *non-separable.* Masalah optimasi sebelumnnya baik pada fungsi obyektif maupun kendala dimodifikasi dengan mengikuti slack variabel $\xi>0$ yang merupkan sebuah ukuran kesalahan klasifikasi. Berikut ini merupakan *constraint* yang sudah dimodifikasi untuk kasus *non separable*:

$$y_i\left[\left(\mathbf{w}^T \cdot \mathbf{x}_i\right) + b\right] \geq 1 - \xi, i = 1,2,...,l \quad (12)$$

Hyperplane atau pemisah yang optimal ditentukan dengan vektor **w**, yaitu dengan meminimumkan fungsi:

$$\Phi(\mathbf{w},\xi) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}\xi_i \quad (13)$$

Dimana C adalah parameter regulasi yang digunakan untuk mengontrol hubungan antara variabel slack dengan $\|\mathbf{w}\|^2$. Bentuk dual dari masalah lagrange menjadi:

$$\max_{\alpha}\mathbf{w}(\alpha) = \max(-\frac{1}{2}\sum_{i,j=1}^{l}\alpha_i\alpha_j y_i y_j\left(\mathbf{x}_i.\mathbf{x}_j\right) + \sum_{i=1}^{l}\alpha_i) \quad (14)$$

Dengan kendala

$$0 \leq \alpha_i \leq C, \quad i = 1,2,...\,l \quad (15)$$

$$\sum_{i=1}^{l}\alpha_i y_i = 0$$

Dan solusi dari masalah ini:

$$\mathbf{w} = \sum_{i=1}^{l}\hat{\alpha}_i y_i \mathbf{x}_i \quad \text{dan} \quad \hat{b} = -\frac{1}{2}\mathbf{w}\left(\mathbf{x}_r + \mathbf{x}_s\right) \quad (16)$$

SVM juga bekerja pada data non linear. Pada dasarnya klasifikasi data non linear memiliki optimasi persamaan:

$$\overline{\alpha} = \arg\min\frac{1}{2}\sum_{i,j=1}^{l}\alpha_i\alpha_j y_i y_j K\left(\mathbf{x}_i,\mathbf{x}_j\right) + \sum_{i=1}^{l}\alpha_i \quad (17)$$

Nilai K(x,y) merupakan fungsi kernel yang menunjukkan pemetaan non linear pada *feature space*. Persamaan ini memberikan *hard classifier* pada pemisahan hyperplane di *feature space*, dengan persamaan:

$$f(\mathbf{x}) = \text{sign}\left(\sum_{SVs}\alpha_i y_i K\left(\mathbf{x}_i,\mathbf{x}\right) + \hat{\mathbf{b}}\right) \quad (18)$$

Dengan nilai :

$$\mathbf{w}^T \cdot \mathbf{x} = \sum_{SVs}\hat{\alpha}_i y_i K\left(\mathbf{x}_i,\mathbf{x}\right) \text{ dan}$$

$$b = -\frac{1}{2}\sum\overline{\alpha}_i y_i\left[K\left(\mathbf{x}_r,\mathbf{x}_i\right) + K\left(\mathbf{x}_s,\mathbf{x}_i\right)\right] \quad (19)$$

Untuk fungsi kernel yang dapat dipakai pada SVM adalah :

1. *Polynomial:* $(\boldsymbol{x^T x_i}+1)^p$
2. *Radial basic function (RBF):* $\exp\left(-\frac{1}{2\sigma^2}||\boldsymbol{x}-\boldsymbol{x_i}||^2\right)$

C. SVM Multiclass

Pada kasus SVM *multiclass* dapat menggunakan beberapa metode yaitu satu lawan semua (SLA), satu lawan satu (SLU) dan *one optimization problem*. Metode yang digunakan pada penelitian ini adalah satu lawan semua (SLA). Pada metode ini untuk klasifikasi k-kelas, menemukan k fungsi pemisah, dimana k adalah banyaknya kelas.

Misalkan fungsi pemisah disebut ρ. Dalam metode ini ρ^i ditrain dengan semua data dari kelas-i dengan label +1 dan semua data dari kelas lain dengan label -1. Jika kita memiliki λ data untuk training $(x_1,y_1),\ldots,(x_i,y_i)$ dimana $\mathbf{x_i} \in R^n$, i= 1,2,.., λ adalah data input dan $y_i \in S=\{1,\ldots,k\}$ kelas dari x_i yang bersangkutan, fungsi pemisah ke-i menyelesaikan persoalan optimasi berikut:

$$\min_{w^i,b^i,\xi_j^i}\{\frac{1}{2}(w^i)^T w^i + C\sum_{j=1}^{\lambda}\xi_j^i\} \quad (20)$$

Setelah menyelesaikan persamaan 20, ada *k* fungsi pemisah yaitu $w^1x+b^1, w^2x+b^2, \ldots, w^kx+b^k$. Kemudian kelas dari suatu data/obyek baru x ditentukan berdasarkan nilai terbesar dari fungsi pemisah j= kelas x= arg $\max_{i=1,..,k} w^ix+b^i$, dimana $j \in S$

D. Breast Cancer

Breast cancer merupakan kanker tertinggi yang diderita perempuan Indonesia yaitu 26 per 100.000 perempuan [2]. Pemeriksaan payudara sejak dini berguna untuk memastikan bahwa payudara seseorang masih normal. Bila ada kelainan seperti infeksi, tumor, atau kanker dapatditemuakan lebih awal. *Breast cancer* yang diobati pada stadium dini kemungkinan sembuh mendekati 90%. Secara keseluruhan faktor yang berpengaruh pada stadium dibagi menjadi tiga bagian yaitu ukuran tumor (T) yang dapat dilihat dari ukuran diameter tumor, yang kedua nodus (N) yang berhubungan dengan keadaan metastase kalenjer aksila, dan yang ketiga adalah metastase (M) yang menunjukkan atau tidak metastase. Usia merupakan salah satu faktor resiko *breast cancer.* Pada hasil mamografi malignant dan benign merupakan faktor yang menunjukkan jinak atau tidaknya *breast cancer*. Disamping itu pada penelitian Yuh-jye lee (2011) menerangkkan bahwa *chemoterapi* sangat menentukan sembuh atau tidaknya *breast cancer* [9].

Dalam penentuan tingkat keganasan dapat dilihat dengan stadium penderita *breast cancer* yaitu Stadium I dengan peluang untuk hidup dalam waktu 5 tahun sebesar 87%, stadium II peluang untuk hidup dalam waktu 5 tahun sebesar 75%, dan stadium III peluang untuk hidup dalam waktu 5 tahun sebesar 46%.

III. METODOLOGI PENELITIAN

A. Sumber Data dan Variabel Penelitian

Sumber data yang digunakan penelitian ini adalah data pasien penderita *breast cancer* di Rumah Sakit "X" pada tahun 2011 berjumlah 178. Data ini merupakan data pasien yang telah melakukan *biopsy.*

Variabel yang digunakan dalam penelitian ini terdiri atas variabel respon (Y) dan variabel prediktor (X).Variabel respon yang digunakan adalah kategori jenis stadium penderita *breast cancer*, yang terdiri dari 3 kategori yaitu:

Y = (1) Stadium I
Y = (2) Stadium II
Y = (3) Stadium III

Pada penentuan variabel prediktor, terlebih dahulu dilakukan konsultasi dengan dokter dan bidang riset di rumah sakit "X" sehingga didapatkan beberapa variabel prediktor pada tabel 1. Variabel-variabel yang didapatkan pada penelitian ini, merupakan hasil dari *biopsy* pasien penderita *breast cancer*. Berikut ini merupakan variabel predictor tingkat keganasan *breast cancer*. Variabel yang tercantum dibawah ini sudah tercantum di form *biopsy*.

Tabel 1.
Variabel Prediktor Penelitian

Variabel	Kategori
Ukuran Tumor T (X_1)	0=T1; 1=T2; 2=T3; 3=T4
Nodus N(X_2)	0=N0; 1=N1; 2=N2; 3=N3;
Cemotheraphi (X_3)	(0) = Chemoterapi (1) =Nochemoterapi
Malignant/Benign (X_4)	(0)= Malignant; (1)=Benign
Letak Kanker (X_5)	(0)=Left/Kiri; (1)=Right/Kanan
Usia Pasien (X_6)	0=Lebih dari 30 tahun 1==kurang dari 30 tahun

B. *Langkah Analisis*

Dalam melakukan penelitian harus dilakukan analisis yang tepat. Berikut ini merupakan langkah-langkah penelitian:

1. Melakukan pengumpulan data sekunder penderita *breast cancer* di rumah saikit "X" sesuai variabel prediktor dan variabel respon.
   a. Menerjemahkan variabel dari bahasa medis menjadi variabel pada tabel 1, dengan dokter dan tim riset rumah sakit
   b. Melakukan pengkodingan data
2. Melakukan interpretasi statistika deskriptif pasien penderita *breast cancer.*
3. Melakukan klasifikasi tingkat keganasan penderita *breast cancer* dengan menggunakan analisis regresi logistik ordinal:
   a. Melakukan estimasi parameter
   b. Melakukan pengujian parameter secara serentak dan individu untuk mengetahui variabel berpengaruh dalam model.
   c. Membuat model logit
   d. Melakukan pengujian kesesuaian model yang telah diperoleh.
   e. Menghitung ketepatan klasifikasi regresi logistik ordinal
4. Melakukan klasifikasi tingkat keganasan penderita breast cancer dengan metode Support Vektor Machine. Berikut ini merupakan algoritma metode SVM:
   a. Melakukan transformasi data sesuai dengan metode SVM multi kelas.
   b. Menentukan fungsi pemisah dengan metode multi kelas satu lawan banyak (SLA).
   c. Menentukan nilai-nilai parameter C= 10, 100, dan 1000, sekaligus menentukan fungsi kernel RBF dengan σ =1, 2, 3 dan fungsi kernel polynomial dengan p=1, 2, 3
   d. Memilih nilai parameter C terbaik.
   e. Menghitung nilai ketepatan klasifikasi terbaik
5. Melakukan perbandingan ketepatan klasifikasi antara analisis regresi logistik ordinal dan *SVM* .

IV. HASIL DAN PEMBAHASAN

A. *Pemodelan Breast Cancer Menggunakan Regresi Logistik Ordinal*

Pasien *breast cancer* pada penelitian ini berjumlah 178. Rata-rata usia pasien *breast cancer* adalah 51 tahun. Pasien yang mengalami stadium I sebanyak 6%, stadium II sebanyak 41% dan staium III sebanyak 53%. Pada analisis dengan menggunakan regeresi logistik ordinal, terlebih dahulu dilakukan uji serentak dan uji individu semua variabel prediktor yaitu ukuran tumor (X_1), Nodus (X_2), *Chemoterapy* (X_3), Malignant/benign (X_4), letak Kanker (X_5), dan Usia pasien (X_6). Hasil dari uji ini akan diadapatkan variabel yang siginifikant.

Berdasarkan uji signifikansi serentak dan individu, diadapatkan variabel yang siginifikant yaitu variabel X_1, X_2, dan X_4. Variabel-variabel ini akan digunakan untuk membentuk model akhir regresi logistik ordinal secara serentak.

Analisis regresi logistik secara serentak terlihat pada Tabel 2. Adapun uji hipotesis adalah sebagai berikut:

H_0 : $\beta_1 = \beta_2 = ... = \beta_6 = 0$

H_1 : minimal ada satu $\beta_k \neq 0$; k = 1, 2, ..., 6

Daerah kritis:

Tolak H_0 jika nilai G > $\chi^2_{(5\%,7)}$ = (14.067)

Tabel 2.
Uji Serentak

Model	G	Chi-Square	Df
Final	39.129	136.622	7

Berdasarkan Tabel 2 menunjukkan bahwa nilai *G* sebesar 39.129 yang berarti nilai G> $\chi^2_{(\alpha,df)}$ kesimpulan yang dihasilkan tolak H_0. Hal ini menunjukkan bahwa variabel bebas ukuran tumor, nodus, dan *malignant/benign* paling tidak ada satu yang berpengaruh secara signifikan terhadap tingkat keganasan *breast cancer.* Langkah selanjutnya dilihat secara parsial signifikansi variabel-variabel tersebut:

Tabel 3.
Uji Parsial Regresi Logistik Ordinal Variabel signifikan

Variabel	Coef	Wald	Odds ratio	Keputusan
Const(1)	-30.146	190.026		Tolak H_0
Const(2)	-23.291	221.699		Tolak H_0
$X_{1(0)}$	-8.585	22.922	0.00018	Tolak H_0
$X_{1(1)}$	-3.813	8.386	0.0220	Tolak H_0
$X_{1(2)}$	-1.1772	1.805	0.308	**Gagal Tolak H_0**
$X_{2(0)}$	-22.448	431.105	$1.78x10^{-10}$	Tolak H_0
$X_{2(1)}$	-20.171	400.85	$1.73x10^{-9}$	Tolak H_0
$X_{4(0)}$	-1.913	0.729	0.14	**Gagal Tolak H_0**

Berdasarkan hasil pada Tabel 3 menunjukkan variabel-variabel yang signifikan. Hasil pada Tabel tersebut juga menunjukkan bahwa variabel X_1 (ukuran tumor) pada kategori

kedua tidak signifikan. Sehingga model logit yang dihasilkan adalah.

Logit 1 : $\hat{g}_1$ (x) = -30.146 – $8.585X_{1(0)}$ - $3.813X_{1(1)}$ – $22.448X_{2(0)}$ - $20.171X_{2(1)}$

Logit 2 : $\hat{g}_2$ (x) = -23.291- $8.585X_{1(0)}$ - $3.813X_{1(1)}$ – $22.448X_{2(0)}$ -$20.171X_{2(1)}$

Logit 1 untuk penderita *breast cancer* yang menderita stadium I dan Logit 2 untuk penderita *breast cancer* yang menderita stadium II dan stadium I.

Berdasarkan logit diatas dapat diketahui bahwa nilai odd ratio sebesar 0.00018, hal ini menunjukkan bahwa pasien *breast cancer* yang ukuran tumornya berada pada T_1 peluang mengalami stadium I lebih kecil jika dibandingkan dengan pasien dengan ukuran tumor bertipe T_4. Sedangkan *breast cancer* dengan ukuran tumor T_2 memiliki peluang mengalami stadium I lebih kecil jika dibandingkan tumor bertipe T_4,.

Sedangkan untuk penderita *breast cancer* yang nodusnya N_0 memiliki peluang mengalami stadium I lebih kecil dibandingkan nilai nodus N_2. Penderita *breast cancer* yang nodusnya N_1 memiliki peluang mengalami stadium I lebih kecil dibandingkan nilai nodus N_2

Untuk nilai peluang pada pasien pertama, dengan menggu-nakan fungsi logit dan fungsi peluang dengan rumus diatas didapatkan nilai $\hat{\pi}_1(x)$=1.43 x 10^{-23}, $\hat{\pi}_2(x)$=1.22 x 10^{-20}, dan nilai $\hat{\pi}_3(x)$=0.999. Nilai tersebut menunjukkan bahwa untuk pasien pertama, mengalami *breast cancer* stadium I memiliki peluang sebesar 1.43 x 10^{-23}, sedangkan peluang menderita breast cancer stadium III sebesar 0.99. Berdasarkan nilai tersebut menunjukkan bahwa peluang tertinggi pasien pertama akan menderita *breast cancer* stadium III.

Berdassarkan uji, model tersebut sudah sesuai dengan nilai P-value sebesar 0.37> α (0.05). Berikut ini merupakan ketepatan klasifikasi dengan menggunakan regresi logistik ordinal untuk data testing

Tabel 4.
Ketepatan klasifikasi

Data original	Prediksi			Ketepatan klasifikasi
	Stadium I	Stadium II	Stadium III	
Stadium I	0	0	4	56.6%
Stadium II	0	0	19	
Stadium III	0	0	30	

Berdasarkan Tabel 4.9 menunjukkan bahwa semua data deviance stadium I dan stadium II diprediksi salah ke stadium III, sedangkan stadium II diprediksi benar stadium III sebanyak 30 data. Secara keseluruhan hasil dari prediksi dengan mengguna-kan regresi logistik ordinal, dimana data training yang digunakan sebanyak 125 dan data testing yang digunakan sebanyak 53 menghasilkan ketepatan klasifikasi sebesar 56.60%.

B. Klasifikasi Tingkat Keganasan Breast Cancer Dengan Menggunakan Support Vector Machine

Analisis SVM pada tingkat keganasan *breast cancer* meng-gunakan dua fungsi kernel yaitu *Polynomial* dan *Radial Basis Function.* Pada fungsi polynomial menggunakan parameter p sebanyak 3 jenis yaitu p=1, p=2, dan p=3. Sedangkan pada *radial basis function* σ =1, σ =2, dan σ =3. Untuk nilai C yang dibandingkan yaitu C=10, C=100, dan C=1000.`Pada analisis tingkat keganasan breast cancer ini menggunakan nilai C yang berbeda. Dengan memasukkan beberapa nilai p pada fungsi kernel menggunakan polynomial dan beberapa nilai σ pada fungsi kernel radial basis function (rbf). Sehingga dapat dibandingkan ketepatan klasifikasi terbaik pada analisis tingkat keganasan breast cancer**.**

Tabel 5
Tingkat Akurasi Klasifikasi SVM Dengan C=10

Proporsi	Polynomial			RBF		
	P=1	P=2	P=3	σ =1	σ =2	σ =3
50:50	81.11	80.89	91.11	96.66	85.39	85.39
70:30	86.71	94.33	98.11	98.11	92.45	88.67
80:20	86.11	94.44	97.22	97.22	91.67	91.67

Berdasarkan Tabel 5 menunjukkan bahwa nilai akurasi tertinggi terdapat pada pengelompokan SVM menggunakan data Training dan Testing 70:30, dengan fungsi kernel yang dipakai polynomial dengan nilai p=3 dan fungsi kernel radial basis function dengan nilai σ =1. Pada pengelompokan ini tingkat akurasi mencapai 98.11 %.

Tabel 6
Tingkat Akurasi Klasifikasi SVM Dengan C=100

Proporsi	Polynomial			RBF		
	P=1	P=2	P=3	σ =1	σ =2	σ =3
50:50	81.11	88.76	91.11	96.66	93.25	91.01
70:30	86.71	96.23	98.11	98.11	98.11	92.45
80:20	86.11	94.44	97.22	97.22	94.44	91.67

Tabel 6 menunjukkan bahwa nilai akurasi tertinggi pada C=100 sama dengan pengelompokan C=10 yaitu akurasi tertinggi terdapat pada pengelompokan SVM mengguna-kan data Training dan Testing 70:30, dengan fungsi kernel yang dipakai polynomial dengan nilai p=3, fungsi kernel radial basis function dengan nilai σ =1 dan nilai σ =2 . Pada kedua jenis pengelompokan ini tingkat akurasi mencapai 98.11 %.

Tabel 7

Proporsi	Polynomial			RBF		
	P=1	P=2	P=3	σ =1	σ =2	σ =3
50:50	88.76	88.76	92.13	96.62	92.13	92.13
70:30	86.79	96.22	98.11	98.11	98.11	98.11
80:20	86.11	94.44	97.22	97.22	97.22	97.22

Tingkat Akurasi Klasifikasi SVM Dengan C=1000

Tabel 7 menunjukkan bahwa nilai akurasi tertinggi pada C=1000 sama dengan pengelompokan C=10 yaitu akurasi terting-gi terdapat pada pengelompokan SVM menggunakan data Training dan Testing 70:30, dengan fungsi kernel yang dipakai polynomial dengan nilai p=3, fungsi kernel radial basis function dengan nilai σ =1, σ =2, σ =3. Keempat jenis pengelompokan ini tingkat akurasinya sama yaitu mencapai 98.11 %.

Selanjutnya membandingkan metode regresi logistik ordinal dan metode Support Vector Machine. Nilai prediksi tingkat keganasan *breast cancer* pada regresi logistik

didapatkan dari nilai peluang tertinggi pada masing-masing kategori. Dan klasifikasi pada *support vector machine* (SVM) diperoleh dari hyperplane yang memisahkan kategori tingkat keganasan *breast cancer.*

Tabel 8.
Perbandingan klasifikasi regresi logistik ordinal dan SVM menggunakan Polynomial p=1,2,3 dan rbf σ =1,2,3

Prop.	Regresi Log. Ordinal	Polynomial dengan P=1,2,3			RBF dengan σ =1, 2, 3		
		1	2	3	1	2	3
50:50	56.60	83.6	86.1	91.4	96.6	90.3	89.5
70:30	53.90	86.7	95.5	98.1	98.1	96.2	93.0
80:20	52.77	86.1	94.4	97.2	97.2	94.3	93.5

Tabel 8 menunjukkan bahwa ketepatan klasifikasi dengan menggunakan *support vector machine* (SVM) lebih baik jika dibandingkan dengan regresi logistik ordinal. Hal ini bisa dilihat dari nilai ketepatan klasifikasi *breast cancer,* pada regresi logistik ordinal ketepatan klasifikasi tertinggi 56.60% dengan proporsi data training dan testing 70:30. Untuk SVM ketepatan klasifikasi rata-rata tertinggi mencapai 98.1% yaitu pada data training dan testing 70:30 dengan menggunakan kernel polynomial p=1 dan RBF $\sigma = 1$.

V. KESIMPULAN/ RINGKASAN

Berdasarkan hasil klasifikasi tingkat keganasan *breast cancer* dengan menggunakan regersi logistik ordinal dan *Support Vector Machine* (SVM) menghasilkan kesimpulan bahwa Model Logit dari regresi logistik ordinal adalah $\hat{g}_1$ (x) = -30.146 – $8.585X_{1(0)}$ - $3.813X_{1(1)}$ – $22.448X_{2(0)}$ - $20.171X_{2(1)}$ dan $\hat{g}_2$ (x) = -23.291- $8.585X_{1(0)}$ - $3.813X_{1(1)}$ – $22.448X_{2(0)}$ - $20.171X_{2(1)}$. Dari model tersebut didapatkan ketepatan klasifikasi tingkat keganasan breast cancer tertinggi 56.60%. Klasifikasi dengan menggunakan SVM menggunakan fungsi kernel radial basis function (RBF) dan Polynomial menghasilkan ketepatan klasifikasi tertinggi mencapai 98.11%. Hasil tersebut menunjukkan bahwa SVM memiliki ketepatan klasifikasi lebih baik jika dibandingkan dengan regresi logistik ordinal

DAFTAR PUSTAKA

[1] WHO (2005). Data penderita breast cancer di dunia [Online]. Available: http://www.who.int/cancer/ detection/breastcancer/en/index1.html.

[2] Dinas Kesehatan Nasional (2007). Data penderita breast cancer di Indonesia [Online]. Available: http://www.depkes.go.id/index.php/berita/press-release/1060-jika-tidak-dikendalikan-26-juta-orang-di-dunia-menderita-kanker-.html

[3] Tim dokter RS Onkologi, *Breast Phsician Course*, Surabaya: RS Onkologi Surabaya (2003).

[4] Ilias Maglogiannis, Elias Zafiropolous, dan Ioannis Anagnostopoulos, **"**An Intelligent System for Automated Breast Cancer Diagnosis and Prognosis Using SVM Based Classifiers," *Applied Intelligence*, Vol. 30, No. 1 (2009) 24–36.

[5] Hui-Ling Chen, Bo Yang, Jie Liu, dan Da-You Liu, "A support vektor machine Classifier with rough set-based feature selection for breast cancer diagnosis," *Expert Systems with Applications*, Vol. 38, No. 7 (2011) 9014-9022.

[6] D. W. Hosmer dan Lemenshow, *Applied Logistik Regression*, USA: John Wiley and Sons (2000).

[7] V. N. Vapnik, *The Nature of Statistical Learning Theory*, New York: Springer (1999).

[8] A. S. Nugroho, A. B. Witarto, dan D Handoko (2003). Support Vektor Machine –Teori dan Aplikasi dalam Bioinformatika [Online]. Available: http://www.Ilmukomputer.com.

[9] Yuh-Jyee Lee, *Support Vektor machines in data mining*, Madison**:** University of Wisconsin (2001).